Monte Carlo Estimation of Accuracy Gains from Option and Pair Elimination in Matching Tasks

# Abstract

Matching word problems are common in tests and many test-taking strategies have been developed, both explicitly and implicitly, to improve chances of correctly answering them. Unlike other variations of multiple-choice questions, matching word questions are rather difficult to guess correctly. Matching word problems usually consist of multiple terms on the left, and multiple definitions or answers on the right, usually unordered (Figure 0). The test-taker's goal is to match them one-to-one correctly.

In this exploration, we seek to measure the gains in accuracy from two common strategies in matching word problems: eliminating possible matches for a given term and specific pairings. Both are simulated based on the additional premise that the removal is always of an incorrect term or pairing, as test-takers sometimes eliminate the correct option.

A paper with text on it

AI-generated content may be incorrect.

Figure 0: an example GRE word-definition matching problem.

# Introduction

A matching word problem will always have the same amount of choices on the left and right, denoted by . Thus, the total possible amount of permutations for matching terms to n definitions is given as

The expectation of the number of correct answers (matches) for this type of problem, when guessing randomly, is always 1, irrespective of n. This can be derived through the linearity of expectation:

This was reduced to 1 via linearity of expectation.

This paper explores how the expected number of correct matches changes when test-takers employ elimination strategies. We categorize our analysis into two primary scenarios:

1. **Correct Answers May Not Be Eliminated (Guaranteed Incorrect Elimination):** This scenario assumes the test-taker perfectly identifies and eliminates only incorrect options or pairs. This represents an ideal, highly effective elimination strategy.
2. **Correct Answers May Be Eliminated (Random Elimination):** This scenario models a less precise elimination strategy where the test-taker randomly eliminates options or pairs, potentially removing correct matches.

# Methods

Our approach utilizes Monte Carlo simulations to estimate the expected number of correct matches under various elimination strategies. The problem is represented as an matrix, where rows correspond to terms and columns to definitions. Initially, all entries in this matrix are true, indicating a possible match. Eliminated options or pairs are set to false.

The simulations were conducted for n values ranging from 1 to 6 (maxN = 6), with each data point averaged over 5000 (trials = 5000) random permutations.

## Elimination Strategies

We investigated the following elimination strategies:

1. **Random Guessing (Baseline):** For each , random permutations are generated, and the number of correct matches is averaged. This serves as the baseline for comparison.
2. **Eliminate Wrong Options per Word (Guaranteed Correct):** For each word (row i), *incorrect* definitions are randomly selected from the available incorrect definitions and marked as false. This ensures that the correct match for word is never eliminated. The parameter k varied from 1 to 3 (k\_list = 1:3).
3. **Eliminate Incorrect Global Pairs (Guaranteed Correct):** A total of *incorrect* word-definition pairs are randomly selected from the entire set of incorrect pairs and marked as false. This is a global elimination strategy where the eliminations are not tied to a specific word. The parameter ranged from 1 up to (i.e., up to 18 for , ) (y\_list = 1:max\_y\_needed).
4. **Eliminate Options per Word (May Forbid Correct):** For each word (row ), definitions are randomly selected from all definitions and marked as false. In this scenario, there is no guarantee that the correct match for word will be preserved. This models a less ideal elimination process.
5. **Eliminate Global Pairs (May Forbid Correct):** A total of pairs are randomly selected from all possible word-definition pairs (including correct ones) and marked as false. This is a global elimination strategy where correct matches may inadvertently be removed.

After applying elimination rules, the allowed matrix defines the remaining valid matches. The allValidPerms helper function generates all possible permutations that adhere to these allowed constraints. If no valid permutations exist, that trial is skipped.

## Accuracy Metric

To quantify the effectiveness of each strategy, we used a metric judging improvement towards a perfect score, calculated for each data point (, , or ) on the graphs. This metric measures how much closer a strategy gets to a perfect score ( correct matches) relative to the random guessing baseline.

Where:

* is the expected number of correct matches for a given strategy at words.
* is the expected number of correct matches for random guessing at words (which is always 1).
* is the total number of words (and the maximum possible correct matches).

This metric normalizes the gain by the maximum possible gain achievable beyond random chance, providing a more scalable percentage of improvement especially with lower baseline accuracies at high values.

# Results

The simulations provide insights into how different elimination strategies impact the expected number of correct matches. The results are visualized in the following figures:

A graph showing the effect of a word

AI-generated content may be incorrect.

**Figure 1: Effect of Per-Word Elimination on Accuracy (Guaranteed Correct)** This figure (Figure 1) illustrates the expected number of correct matches when *wrong* options are eliminated per word. As n increases, the baseline random guessing accuracy remains constant at 1. However, for , the expected correct matches generally increase, especially for smaller n. The lines show that even a small number of guaranteed eliminations per word can lead to substantial gains. The improvement percentages indicated on the right quantify this gain relative to the potential improvement from random guessing to a perfect score.

A graph showing the effect of global pairs

AI-generated content may be incorrect.

**Figure 2: Effect of Global Pair Elimination on Accuracy (Guaranteed Correct)** Figure 2 presents the results for eliminating *incorrect* pairs globally. Similar to per-word elimination, increasing y generally leads to higher expected correct matches. For small , eliminating even a few pairs can significantly boost accuracy. The curves show a diminishing return as y increases beyond a certain point for a given , as fewer beneficial incorrect pairs remain to be eliminated.

A graph showing the effect of a word

AI-generated content may be incorrect.

**Figure 3: Effect of Per-Word Elimination (May Forbid Correct)** In Figure 3, we observe the impact of per-word elimination where correct options *may* be forbidden. Compared to Figure 1, the accuracy gains are significantly diminished and more erratic. Randomly eliminating options, even if only a few, introduces a risk of removing the correct answer, which can severely hinder performance. For larger , the expected matches tend to hover closer to the random baseline, and in some cases, can even fall below it if too many correct options are inadvertently removed.

A graph showing the effect of global palm elimination

AI-generated content may be incorrect.

**Figure 4: Effect of Global Pair Elimination (May Forbid Correct)** Figure 4 depicts the scenario where pairs are randomly eliminated globally, potentially including correct pairs. Similar to Figure 3, the gains are much less pronounced and more volatile than when only incorrect pairs are eliminated. The risk of removing correct pairs globally can lead to accuracy fluctuations and, for higher y values, can even reduce the expected correct matches below the random guessing level, highlighting the importance of accurate elimination.

A graph showing the number of words

AI-generated content may be incorrect.

**Figure 5: Dynamic Global Elimination** This figure (Figure 5) compares the random baseline against global elimination strategies where the number of eliminated pairs () is dynamically linked to and by the formula . Each line represents a fixed value of k (e.g., , k), but the *absolute number of pairs eliminated* () increases as increases along that line. This illustrates how the total impact of per-word elimination scales with problem size. The curves show that even a constant can lead to significant and varying improvements as n grows, because the total number of eliminated pairs grows proportionally with .

A graph showing the number of words

AI-generated content may be incorrect.

**Figure 6: Global Elimination with Specific Constant Absolute Eliminations** Figure 6 specifically isolates the effect of eliminating a *constant absolute number* of incorrect pairs, independent of n. The lines shown correspond to eliminating 2, 6, and 12 pairs, respectively. These specific y values are chosen because they represent the total number of incorrect pairs in problems of size , , and when all incorrect pairs are eliminated (i.e., ). This graph allows for a direct comparison of the impact of a fixed level of knowledge/elimination across varying problem sizes. It demonstrates that a fixed amount of elimination has a more pronounced effect for smaller (where it represents a larger proportion of total pairs) and a diminishing effect as increases.

# Discussion and Conclusion

Our Monte Carlo simulations provide empirical evidence supporting the intuition that elimination strategies significantly improve accuracy in matching tasks. The distinction between eliminating only incorrect options/pairs and potentially eliminating correct ones is critical. Strategies that guarantee the removal of only incorrect information consistently yield substantial gains, especially for smaller problem sizes (). In contrast, strategies that risk removing correct answers show much weaker and more unpredictable improvements, underscoring the importance of accurate knowledge in test-taking.

The comparison between per-word elimination and global pair elimination, particularly when scaled to equivalent total eliminations (), reveals that the overall impact on accuracy is a complex function of both the number of terms (n) and the amount of information eliminated. The "-equivalent" dynamic elimination graph (Figure 5) highlights that a fixed per-word strategy leads to a growing absolute number of eliminations as n increases, resulting in continued accuracy gains. Conversely, the "constant absolute eliminations" graph (Figure 6) clearly shows that a fixed amount of eliminated knowledge has a greater relative impact on smaller problems.

Future work could involve:

* Investigating analytical solutions for the "May Forbid Correct" scenarios, which are more complex due to the inclusion-exclusion principle.
* Exploring hybrid strategies where test-takers have partial certainty about eliminations.
* Analyzing the impact of different distributions of incorrect answers or test-taker biases.
* Extending n to larger values to observe asymptotic behavior, though this would require more computationally efficient methods than brute-force permutation generation.

This study reinforces the value of effective elimination strategies in matching tasks, emphasizing that precision in elimination is paramount for maximizing accuracy gains.

# Figures

A graph showing the effect of a pair of air

AI-generated content may be incorrect.A graph showing the effect of a word

AI-generated content may be incorrect.A graph of different colored lines

AI-generated content may be incorrect.A graph showing the effect of incorrect per-word

AI-generated content may be incorrect.